American Journal of Medical Genetics 64:319–322 (1996)

Rater Reliability of Fragile X Mutation Size Estimates: A Multilaboratory Analysis

Gene S. Fisch, Nancy Carpenter, Anne Maddalena, Jack Tarleton, Colin Julien-Inalsingh, and Jeanette J.A. Holden

Kings County Hospital Center and SUNY/Health Science Center, Brooklyn, New York (G.S.F.); Chapman Institute of Medical Genetics, Tulsa, Oklahoma (N.C.); Genetics and IVF Institute, Fairfax, and Medical College of Virginia, Richmond, Virginia (A.M.); Greenwood Genetics Center, Greenwood, South Carolina (J.T.); Ongwanda Resource Centre, Kingston, Ontario (C.J.-I., J.J.A.H.); and Queen's University, Kingston, Ontario, Canada (J.J.A.H.)

Notwithstanding the use of comparable molecular protocols, description and measurement of the fra(X) (fragile X) mutation may vary according to its appearance as a discrete band, smear, multiple bands, or mosaic. Estimation of mutation size may also differ from one laboratory to another. We report on the description of an mutation size estimate for a large sample of individuals tested for the fra(X) pre- or full mutation. Of 63 DNA samples evaluated, 45 were identified previously as fra(X) pre- or full mutations. DNA from 18 unaffected individuals was used as control. Genomic DNA was extracted from peripheral blood, and DNA fragments from each of four laboratories were sent to a single center where Southern blots were prepared and hybridized with the pE5.1 probe. Photographs from autoradiographs were returned to each site, and raters blind to the identity of the specimens were asked to evaluate them. Raters' estimates of mutation size compared favorably with a reference test. Intrarater reliability was good to excellent. Variability in mutation size estimates was comparable across band types. Variability in estimates was moderate, and was significantly correlated with absolute mutation size and band type.

KEY WORDS: trinucleotide repeats, reliability, variability, fragile X

INTRODUCTION

A review of the fra(X) (fragile X) literature provides ample evidence of interlaboratory variation in the qualitative description of fra(X) mutations, even among laboratories using the same restriction enzymes and probes. Obvious variables include electrophoretic migration distance, amount of DNA loaded, and geometry of wells, as well as background and exposure time of autoradiographs. A large multicenter study employing a common method (*Eco*RI and *Eag*I Southern blot) presented an opportunity to isolate and study one of these factors, i.e., variation among individual readers performing visual evaluation of mutation size and type.

We report on the description of and mutation size estimate for a large sample of individuals tested for the fra(X) mutation. The purposes of our study were to determine: 1) reliability of mutation size estimates and band type identification from Southern blots; 2) range of variability; and 3) source(s) of variability in estimating size.

MATERIALS AND METHODS

Subjects

Forty-five males and females age 3–15 years, diagnosed with the fra(X) mutation, were tested during the past 4 years. Initial diagnosis was made from DNA testing, using either the StB12.3 or pE5.1 probe and *Eag*I/*Eco*RI double digests [Oberlé et al., 1991]. Thirty-three samples were from males, and 12 were from females. The full mutation was detected in 39 subjects. Six were diagnosed with the premutation. Subjects were obtained and tested at each of four sites: 1) 12 were from the Chapman Institute in Tulsa, OK; 2) 13 were from the Ongwanada Resource Centre in Kingston, Ontario, Canada; 3) 13 were from the Genetics and IVF Institute in Fairfax, VA; and 4) 7 were from the Greenwood Genetics Center in Greenwood, SC.

DNA Testing

Genomic DNA was extracted from blood and digested using *Eag*I and *Eco*RI enzymes. Digested DNA from each location was then sent to a single test site

Received for publication September 15, 1995; revision received January 18, 1996.

Address reprint requests to Gene S. Fisch, Department of Psychiatry, Box 32, Kings County Hospital Center, 451 Clarkson Ave., Brooklyn, NY 11203.

(Ongwanada), where Southern blots were prepared and hybridized with the pE5.1 probe. Lambda *Hind*III and *Eco*RI fragments were included as molecular weight standards, and lambda DNA was included in the hybridization. Photographs (5″ × 7″) of autoradiographs of all subjects were then returned to each testing center, where laboratory staff blind to the patient numbers and diagnoses evaluated all subjects in the study and sized bands in each lane.

Data Analysis

Mutation size estimates from each rater were separated into premutations and full mutations. Full mutations were further subdivided by type: a) discrete bands; b) multiple bands; and c) smears. Raters were also asked to identify samples that appeared to be methylation mosaics. For discrete-band full mutations and premutations, raters' estimates were ranked from lowest to highest using the median to establish size. For multiple bands, size was represented by the band with the strongest signal. Median estimates were obtained using the ranking method described. For smears, the logarithmic average of the top and bottom estimates of the spread was calculated for each rater. Log averages were then ranked, and the median value was chosen.

To measure variability in the raters' estimates, the difference between the highest and lowest mutation size estimates was calculated for each subject and for all samples.

Statistical Analysis

An important property of the measurement process, its reliability, is obtained by evaluating measurement error, and is defined as the ratio of true variation to observed variance [Dunn, 1989]. Calculations of reliability are contingent upon the measurement level of the instrument. We computed the Pearson correlation coefficient and intraclass correlation, both of which were obtained from a one-way analysis of variance. We also evaluated measurement error by determining the coefficient of variation.

Rater reliability for mutation-size estimates were assessed in two ways. The first compared raters' estimates to those from a reference test, i.e., interrater reliability. The second measure compared raters' estimates for subjects whose DNA were run twice albeit in different lanes, i.e., intrarater reliability. Internal and external validity of the reference test was obtained from the coefficient of variation.

Internal and External Validity of the Reference Test

For the reference test, one rater employed a digitizer (Science Accessory Corp. model MT1 running Compugene software, Compugene Inc, Chesterfield, MO). To determine internal validity for this study, a subsample (3 fra(X) males, 3 fra(X) females, 1 control) was measured repeatedly. Previously, internal validity for the digitizer was calibrated using placental DNA of known sizes ranging from 2–9 kb. External validity was established using measured band sizes which were compared frequently with those calculated from an automated image analyzer.

Intrarater Reliability

Intrarater reliability estimates were computed for 19 subjects whose DNA were run twice. Intrarater reliabilities for three raters and the reference test were calculated from intraclass correlation coefficients.

RESULTS

From the subsample evaluated, internal validity for this study showed a mean coefficient of variation of 0.63%, indicating a reliability of 99.37%. Previously, the coefficient of variation for the digitizer using placental DNA was <0.9%. To compare interrater reliability for individual raters, we computed separate regression analyses, which paired each of their estimates with that of the reference test [Dunn, 1989]. Pearson correlation coefficients, shown in Table I, indicate strong, significant relationships between each rater's estimate and that of the digitizer. Intraclass correlations in Table I indicate significant moderate-to-strong intrarater reliability.

To obtain interrater reliability between the three raters (as a group) with the reference test, we computed the arithmetic mean of their estimates with the size ascertained by digitizer. Correlation between reference test and mean rater estimate was $r = .94$ ($P < .0001$). Results are shown in Figure 1.

Agreement according to band-type description (discrete, smear, multiple, or mosaic) was more difficult to assess. For the 19 retested subjects (38 lanes, or 19 pairs), all three raters agreed to a specific band type on 13 samples, while 2/3 raters agreed on the remaining specimens. Agreement was higher for samples identified as mosaics. Of the 7 pairs so classified, all three raters agreed that 6 pairs were mosaics.

Mutation Size Estimates

We examined variability, i.e., highest-to-lowest difference, in size estimates, mutation type, and band type to determine if there were any obvious relations among them. For each band type, we computed the mean high-low difference in raters' estimates for males and females. Results in Table II show that regardless of band type or sex, mean differences in full mutation size estimates among the four raters were comparable. Premutations are inherently smaller, and hence differences in their estimates were narrow. A more detailed

TABLE I. Correlation Coefficients for Calculating Rater Reliability by Comparing Raters' Mutation Size Estimates to a Reference Test, and by Intrarater Reliability*

	Reference test (digitizer)	Intrarater reliability
Reference test		.99 ($P < .01$)
Rater 1	.93 ($P < .001$)	.88 ($P < .01$)
Rater 2	.93 ($P < .001$)	.99 ($P < .01$)
Rater 3	.90 ($P < .001$)	.63 ($P < .01$)

* Significance levels are in parentheses.

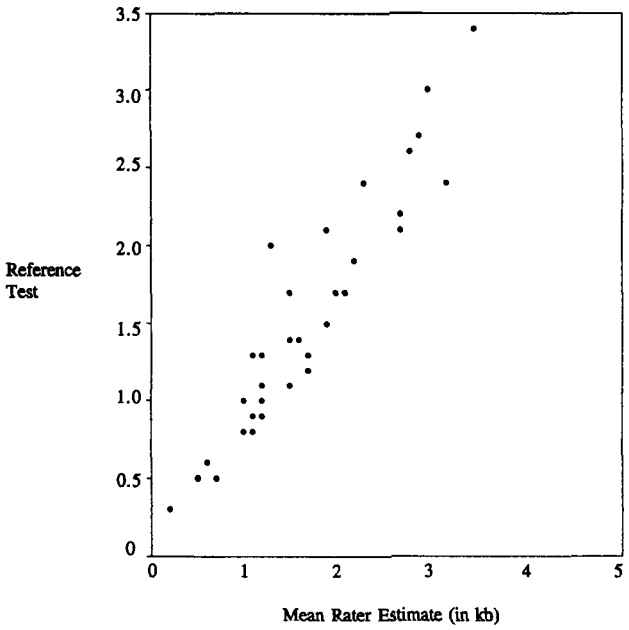

Fig. 1. Mean full mutation size estimate for three raters compared with mutation size estimate ascertained by digitizer (in kb).

examination revealed that differences in highest-to-lowest estimates ranged from 0.15–1.20 kb. Overall mean difference between highest and lowest estimates was 0.47 kb. Table III shows that 62% of differences in estimates ranged from 0.25–0.50 kb.

To determine which variables were most important in producing differences in estimates for full mutations, we employed a multiple regression analysis to examine three factors: absolute mutation size, band type, and sex. Premutations were excluded from analysis, since smaller mutations migrate to parts of the gel where precision is greater. Results show that the correlation of these variables with variability in mutation size estimate was moderate (multiple R = .48) and highly significant ($P < .009$). Absolute mutation size and band type (smears) were both significant factors, but sex was not. Absolute size accounted for the largest proportion of the variability ($r = .38$; $P < .004$).

DISCUSSION

Our analysis indicates that raters' estimates of fra(X) mutation size are highly reliable when compared to a reference test. Intrarater reliability was good to excellent. Raters generally agreed on band type observed.

Variability, i.e., highest-to-lowest difference, in raters' estimates is associated primarily with absolute mutation size. However, absolute size and band type (smears) both contribute to variability in estimates, although smears have a smaller effect. Nonetheless, differences in estimates are modest, given absolute size, and range mostly from 250–500 bp. Variability in estimates is comparable for male and female samples, and stable across band type. However, precision in estimating mutation size will not be critical so long as individuals with full mutations and premutations are identified properly. In that regard, all full-mutation samples examined in this study were categorized correctly. Thus, while variability in estimates is greater for larger mutations, differences in sizing do not affect identification of the full mutation.

Since presence of the FMR-1 protein (or lack thereof) is related to degree of methylation, precision in estimating mutation size is not as crucial as establishing methylation status in the phenotype. Moreover, mutation-size estimates are less important than degree of methylation when considering extent of cognitive deficit [Rousseau et al., 1994; Steyaert et al., 1996; de Vries et al., 1996]. Full mutation size is unrelated to levels of cognitive or adaptive behavior deficits [Fisch et al., 1996a], as well as to longitudinal changes in cognitive or adaptive behavior deficits [Fisch et al., 1996a,b]. Earlier, McConkie-Rosell et al. [1993] and Hagerman et al. [1994] reported that males with incompletely methylated full mutations have milder cognitive deficits.

Sources of variability not examined here were reported by the raters themselves. Not all were familiar with the amount of background or electrophoretic migration distance used, nor were they accustomed to evaluating so many gels at once. These factors probably account for the lower correlations ascertained for intrarater reliability.

Finally, digestion with *Eco*RI and *Eag*I enzymes may not produce optimal estimates of mutation size in carriers. We observed two instances in which two raters classified an individual known previously to carry a premutation as a normal male. That is, there were two false-negatives among the premutation samples, neither of which was produced by the digitizer. This may have resulted from overloading the lane in conjunction with a relatively short electrophoretic distance. Repeat analysis with less DNA or longer migration would have been indicated in a clinical situation. Alternatively,

TABLE III. Relative Frequency Distribution of Highest-to-Lowest Differences in Mutation Size Estimates (in kb) for Fully Mutated Fra(X) Males and Females (N = 39)

Highest-to-lowest difference	Relative frequency (%)
0.25 kb or less	12
0.25–0.50 kb	62
0.50–0.75 kb	14
0.75–1.00 kb	9
>1.0 kb	3

TABLE II. Differences in Highest-to-Lowest Mutation Size Estimates (in kb) for Fra(X) Males and Females

	Mean high-low differences in estimates (±SD)	
Band type	Males	Females
Full mutations		
Discrete	0.54 (0.29)	0.48 (0.26)
Smear	0.54 (0.29)	0.40 (0.18)
Multiple	0.46 (0.19)	N/A
Premutations	0.12 (0.05)	0.20 (0.11)

other complementary methods that detect premutations could be used, such as *Pst*I blots or PCR.

ACKNOWLEDGMENTS

We thank Burhan Say, Director, Chapman Institute of Medical Genetics, Tulsa, OK, Joseph D. Schulman, Director, Genetics and IVF Institute, Fairfax, VA, and Roger Stevenson, Director, Greenwood Genetics Center, Greenwood, SC, for their interest in this project and willingness to participate. We also thank Jane Pritchard, Linda Randell, Susan H. Black, Suzanne Holowinsky, Charles E. Schwartz, Angarad Holmes, Pat Frost, Beverly Lee, and Ruth Lokkesmoe for their invaluable assistance. Studies in Kingston, Ontario, Canada, were supported by a research grant from the Ontario Mental Health Foundation to J.J.A.H.

NOTE ADDED IN PROOF

Reanalysis of two DNA samples identified previously as a premutation made and denoted normal by 2 raters indicates that both samples belong to a FRAXE male.

REFERENCES

De Vries BBA, Wiegers AM, Smits APT, Frijns J-P, Curfs LMG, van Oost BA, Halley DJJ, Oostra BA, van den Ouweland AMW, Niermeijer MF (1996): Mental status and FMR1 gene mutation in females. Am J Med Genet 64:XX–XX.

Dunn G (1989): "Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors." New York: Oxford University Press, pp 1–18, 40–57.

Fisch GS, Carpenter N, Howard-Peebles PN, Maddalena A, Simensen R, Tarleton J, Julien-Inalsingh C, Chalifoux M, Holden JJA (1996a): Lack of association between mutation size and cognitive/behavior deficits in fragile X males: A brief report. Am J Med Genet 64: 360–362.

Fisch GS, Chalifoux M, Holden JJA, Carpenter N, Howard-Peebles PN, Maddalena A, Simensen R, Tarleton J (1996b): Longitudinal study of cognitive abilities and adaptive behavior levels in fragile X males: A prospective multicenter analysis. Am J Med Genet 64: 354–359.

Hagerman RJ, Hull CE, Safanda JF, Carpenter I, Staley LW, O'Conner RA, Seydel C, Mazzocco MMM, Snow K, Thibodeau SN, Kuhl D, Nelson DL, Caskey CT, Taylor AK (1994): High-functioning fragile X males: Demonstration of an unmethylated, fully expanded FMR-1 mutation associated with protein expression. Am J Med Genet 51:298–308.

McConkie-Rosell A, Lachiewicz AM, Spiridigliozzi GA, Tarleton J, Schoenwald S, Phelan MC, Goonewardena P, Ding X, Brown WT (1993): Evidence that methylation of the FMR-1 locus is responsible for variable phenotypic expression of the fragile X syndrome. Am J Hum Genet 53:800–809.

Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL (1991): Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097–1102.

Rousseau F, Heitz D, Tarleton J, MacPherson J, Petterson U, Mathew C, Mornet E, Maddalena A, Schinzel A, Marcos JAG, Schorderet DF, Schaap T, Maccioni L, Travi M, Jacobs PA, Schwartz C, Mandel JL (1994): A multicenter study on genotype-phenotype correlations in fragile X syndrome using direct diagnosis with probe StB123: The first 2,253 cases. Am J Hum Genet 51:225–237.

Steyaert J, Borghgraef M, Fryns JP (1996): Cognitive and behavioral profile in mildly affected young fra-X males. Seventh International Workshop on the Fragile X and X-linked Mental Retardation, Tromso, Norway.